Statement of Work
DNA Sequencing Services
RFQ-RT-12-00160

The Office of Research and Development (ORD) is seeking to take advantage of the advances in next generation sequencing to obtain genetic information for bacteria and microbial communities relevant to microbial water quality and public health. One of the objectives of this research is to better understand adverse outcomes associated with exposure to waterborne pathogens and the role microbial communities associated with them have on their fate and transport. Waterborne pathogens are transmitted via the fecal-oral route. Thus, fecal pollution plays an important role in the impairment of water systems used for recreational activities and drinking water sources. Fecal pollution is measured by the presence of microbial surrogates using primarily culture-based methods. While the use of surrogates continues to be common practice, criticism is mounting on their real value to predict health risks. This is in part due to the lack of information on their fate, transport, and source origin. Microbial survival mechanisms can be impacted by abiotic factors such as environmental physo-chemical conditions and by biotic interactions with other members of the millieu. However, the organism's genetic make-up dictates how it can cope with both biotic and abiotic interactions. Genome analyses can also be used to differentiate between organisms of fecal and non-fecal origin. Unfortunately, very little is known on the genetic diversity and potential of fecal indicator organisms (i.e., environmental strains), as much attention has been given to clinical isolates. To address this issue, scientists in ORD are proposing to generate genomic sequencing databases for indicator bacteria and metagenomic sequencing databases for drinking water samples and environmental water samples with different levels of fecal pollution.

The ultimate goal of this program is to develop a better understanding of the genetic core of indicator organisms and relevant microbial communities and used this information to develop robust molecular methods based on ecologically relevant molecular markers. ORD does not currently have the sequencing capabilities required for the completion of this project in a timely manner. As a consequence, we will require analytical services from a laboratory capable of performing genome and metagenome sequencing for the duration of the contract.

The genomes that will be sequenced for this project will be primarily for indicator bacteria such as fecal coliforms and enterococci, and alternate indicators such as *Bacteroides, Clostridium, Prevotella,* and *Bifidobacterium* species. The metagenomes to be studied will be from waters impacted with high level of pollution and/or from drinking water samples. Specifically, we are seeking the use of next generation sequencing methods to generate these sequencing databases. EPA personnel will perform the DNA extractions using established protocols. Samples will be concentrated in DNAse-free tubes at desirable amounts and stored frozen at -20°C until processed by the contractor. Each tube will contain a specific sample that will be used in sequencing reactions performed by the contractor. The project principal investigator (PI) will contact the contractor as for the need to sequence a specific number of samples. The samples will be placed in appropriate coolers, which will then be shipped to the contractor using overnight delivery.

The primary output of this service will consist of electronic files containing the sequencing results, including FastaQ, electropherograms, etc. Each sequencing reaction should generate a minimum of 6.4 million reads of pair-end 100 reactions. It is expected that the data will be compatible with publicly available sequencing analysis softwares.

The contractor should store the data in a secure site, and provide the data to the PI using a file transfer protocol (FTP) approach. Due to the amount of sequences per sample that is required in this project, the contractor must have high throughput capabilities and must be capable of producing high quality sequence data using next generation sequencing technologies (such as Illumina HiSeq 2000). It is estimated that a minimum of 75 samples will be processed as part of this task.

PERIOD OF PERFORMANCE

September, 1st 2012 thru August, 31st 2013.

OBJECTIVE

The contractor shall conduct next generation sequencing of samples from genomic and metagenomic DNA extracted by EPA personnel, and provide sequencing data in electronic form. EPA will provide a total of 75 different samples. Additional samples can be provided to replace any sample as needed (i.e., due to QA/QC failure).

EPA will provide:

1. DNA extracts for each sample to be sequenced.
2. EPA will provide DNA extracts in the concentration recommended by the contractor.
3. A list in spreadsheet format for each sample as an email attachment.

Contractor shall furnish:

1. Tubes and barcodes to be used in the identification of the samples
2. Cleanup kits, sequencing kits, laboratory equipment (gel electrophoresis, centrifuges, thermal cyclers, etc.), and any related reagents needed to complete the tasks herein.
3. Gel images of the products, annotated with the sample names.
4. Establish a pick up/delivery of sample protocol and cover for these expenses.
4. All .fastaq and genome assembly files.

DESCRIPTION OF TASKS

Task 1. DNA manipulation

The contractor shall perform any further purification or concentration of the DNA following standard protocols. Additionally, contractor shall prepared all the index libraries (one index per sample) required for the completion of the project.

Task 2. Sequencing
In order to obtain enough information to generate genomes or to assemble functional genes in metagenome samples, a minimum of 6.4 million reads is required for each sample. The contractor shall perform sequencing using a pair-end (PE) approach and a minimum of 100 bp per PE. Therefore, the contractor shall provide a minimum of 1.28 Giga bases per sample.

QA/QC
The contractor shall comply with next-generation sequencing QA/QC requirements as established by the instrument's manufacturer. The contractor shall provide a minimum of 85% of high quality (Phred 20) reads from the total number of reads required, and an average length of quality read of 85 bp per each PE. Failure to accomplish this shall require additional reactions. The additional reactions needed to complete the minimum number of bases or reads will be performed by the contractor at no extra charge. Data should be collected under standard procedures and conditions for molecular biology analysis. The personnel conducting these experiments must have training and experience in molecular biology techniques in order to obtain quality data. Personnel must be experienced in sample handling, DNA concentration, purification, and quantification, polymerase chain reaction methodologies, gel electrophoresis methods and next generation sequencing methods.

DELIVERABLES
The contractor needs to perform data assembly for each sequencing dataset. Additionally, the contractor shall develop genome (circular) maps using as reference standards the genomes of closely related species. The reference species to generate the assembly and circular maps will be selected in consultation with EPA personnel. The contractor shall perform gene analyses of phylogenetic and house-keeping genes using available algorithms (e.g., Blast) to confirm the identity of the genome. For metagenomic samples the latter information is not required

As EPA will perform additional bioinformatic analyses, we are requesting all the raw data files, including FastQ files and contigs as well as other files that are normally produced as next generation sequencing runs. Preferentially, the data will be accessed by the Work Assignment Manage via internet or sent on an external drive via airmail. The data will not be shared with a third party unless authorized by the Work Assignment Manage.

The contractor shall maintain a backup of this data for the entire duration of the project. The data shall be compatible with sequencing analysis software or similar software. The contractor shall expect to deliver at least 25% of the total analytical services herein requested every three months.

ACCEPTANCE PROCEDURE FOR DELIVERABLES

1. Deliverables shall be delivered to the Work Assignment Manager designated at contract award. The contractor shall submit the revised deliverable within 6 weeks after receipt of Work Assignment Manager's comments.

2. The Work Assignment Manager will advise the contractor of any problems with deliverable within 3 weeks after receipt of the deliverable. The contractor shall submit the revised deliverable within 6 weeks after receipt of Work Assignment Manager's comments.